



קורס:  
פרקטיקום במדעי הנתונים

נושא הפרוייקט:  
ניתוח וחיזוי מחירי מניות

[קישור לתיקיית הדאטה בענן](#)

מגיש:  
היבה עבד אלכרים

## תוכן עניינים

3	רקע:
3	מה ידוע עליו היסטורית:
3	מה נחקר עד היום:
3	מה מעניין/מיוחד/חשוב בנושא זה:
4	מה שונה בעבודה שלכם ממה שהיה ידוע עד היום:
4	כיצד ביצוע עבודה זו יכול לעזור לאנושות/לחברה:
5	השערות ושיטות מחקר
6	שיטות מחקר
7	נתונים:
9	מניפולציות:
10	דיאגרמות פשוטות:
12	עיבוד הנתונים:
12	השערה 1
13	קישור הגרף להשערת המחקר:
14	השערה 2
16	מסקנה לגבי ההשערה 2:
16	השערה 3
17	קשר להשערה 3:
19	חיזוי:
20	דיון ומסקנות:
21	הרחבות לעבודה
22	סיכום

## רקע:

### תחום המחקר: ניתוח ביצועי מניות ומדדים כלכליים

תחום זה עוסק בניתוח נתוני שוק ההון, במיוחד ביצועי מניות ומדדים כלכליים כמו ה-S&P 500. הניתוח כולל בחינה של מחירי מניות, תשואות, מגמות שוק, והשוואת ביצועים בין מניות שונות ונתוני מדדים. תחום זה מהווה חלק מרכזי בתיאוריה ובפרקטיקה של ניתוח פיננסי והשקעות.

### מה ידוע עליו היסטורית:

#### 1. היסטוריה של שוק ההון:

- שוק ההון המודרני: נוצר בתחילת המאה ה-20 עם הקמת הבורסות המרכזיות וההנחות הראשונות של מניות.
- מדדי מניות: מדדים כמו ה-S&P 500 הוקמו כדי לספק אינדיקציה רחבה על ביצועי השוק כולו.
- תיאוריה מודרנית: תיאוריה כמו תיאורת השוק היעיל (Efficient Market Hypothesis - EMH) פותחה בשנות ה-60 וה-70, השואפת להסביר כיצד מחירי מניות משקפים את כל המידע הזמין.

#### 2. חקר תשואות:

- תשואות יומיות וחודשיות: ניתוח תשואות יומיות וממוצעות חודשיות מאפשר להשוות בין ביצועי מניות שונות ולהעריך את התשואות המשוקללות.
- תיאוריות השקעה: קיימות תיאוריות רבות בתחום ניהול תיקי השקעות כמו תיאוריה של פורטפוליו (Portfolio Theory) ותיאוריה של אופטימיזציה של תיקי השקעות (Capital Asset Pricing Model - CAPM).

### מה נחקר עד היום:

#### 1. ניתוח ביצועים ותחזיות:

- מודלים חיזוי: פותחו מודלים רבים לחיזוי מחירי מניות ותשואות, כולל מודלים סטטיסטיים ואלגוריתמים של למידת מכונה.
- מגמות שוק: ניתוח מגמות שוק כולל הבנת דינמיקות של עליות וירידות מחירים והשפעות של אירועים כלכליים.

#### 2. השפעות כלכליות:

- השפעת כלכלה מקרו-כלכלית: ניתוח השפעות של גורמים כלכליים כמו אינפלציה, ריבית ואירועים פוליטיים על מחירי מניות.

### מה מעניין/מיוחד/חשוב בנושא זה:

#### 1. השפעת השוק על כלכלה אישית:

- השקעות אישיות: כיצד תשואות המניות משפיעות על חיסכון והשקעות אישיות של אנשים פרטיים.
- ניהול תיק השקעות: חשיבות ניתוח מדויק של ביצועי מניות לניהול תיקי השקעות מוצלח.

#### 2. תחזיות והשפעתן:

- תחזיות מדויקות: יכולת לחזות שינויים עתידיים בשוק יכולה להוביל לרווחים משמעותיים והשקעות חכמות יותר.

מה שונה בעבודה שלכם ממה שהיה ידוע עד היום:

1. שילוב נתונים שונים:

- השוואת מניות מול מדדים: העבודה משלבת נתוני מניות עם נתוני מדדים כמו ה-S&P 500, ומביאה לתובנות חדשות על ביצועים יחסיים.
- ניתוח תשואות יומיות וחודשיות: מספקת הבנה מעמיקה על התנהגות המניות לאורך זמן ובפרקי זמן שונים.

2. יישום מודלים ויזואליים:

- ויזואליזציה: הצגת נתונים באמצעות דיאגרמות וגרפים מאפשרת להבין את התנהגות השוק בצורה אינטואיטיבית וברורה יותר.

כיצד ביצוע עבודה זו יכול לעזור לאנושות/לחברה:

1. שיפור ניהול השקעות:

- השקעות חכמות יותר: עבודה זו מספקת כלים לניהול תיקי השקעות בצורה מדויקת ומושכלת יותר.
- תחזיות כלכליות: מסייעת למשקיעים ולמנהלי תיקי השקעות לחזות את מגמות השוק בצורה טובה יותר ולבצע החלטות השקעה מושכלות.

2. הבנה כלכלית רחבה:

- הבנת שוק ההון: מסייעת להבנת הדינמיקה של שוק ההון והשפעות גורמים כלכליים על מחירי המניות.
- השפעה על החלטות כלכליות: מספקת תובנות שיכולות להשפיע על החלטות כלכליות של פרטים, עסקים, ומוסדות כלכליים.

באופן כללי, עבודה זו תורמת לידע הכללי על שוק ההון ומספקת כלים מעשיים למשקיעים ולמנהלי תיקי השקעות, שיכולים לשפר את הביצועים הכלכליים ולהוביל לניהול טוב יותר של כספים.

## השערות ושיטות מחקר

הנה סקירה של השערות אפשריות:

### 1. השערה 1: תשואות יומיות משתנות בין מניות שונות

- השערה: תשואות יומיות של מניות שונות יהיו שונות זו מזו, כאשר מניות של חברות שונות מציגות רמות שונות של תשואות יומיות עקב הבדלים בעסקי החברות, מגמות השוק, ותנודתיות.
- שיטה לבדוק:
  - חישוב תשואות יומיות: חישוב תשואות יומיות לכל מניה על פי שינוי במחיר הסגירה.
  - השוואת תשואות: השוואת התשואות היומיות של מניות שונות באמצעות סטטיסטיקות תיאוריות כמו ממוצעים וסטיית תקן, והשוואה בין התשואות המתקבלות עבור מניות שונות.

### 2. השערה 2: יש קשר בין תשואות מניות לבין ביצועי מדדים כמו S&P 500

- השערה: תשואות המניות יהיו מקושרות לביצועים של מדדים רחבים כמו ה S&P 500-כאשר שינוי במדד ייתכן ויגרום לשינויים גם בתשואות המניות.
- שיטה לבדוק:
  - תשואות חודשיות של מניות ומדדים: חישוב תשואות חודשיות עבור מניות ונתוני מדד.
  - קשר בין משתנים: ביצוע ניתוח קורלציה או רגרסיה כדי לבדוק את הקשר בין תשואות מניות לתשואות המדד לאורך זמן.

### 3. השערה 3: מניות בתחום טכנולוגיה מציגות תשואות גבוהות יותר מאשר מניות בתעשיות מסורתיות

- השערה: מניות בתחום הטכנולוגיה יציעו תשואות גבוהות יותר בהשוואה למניות בתעשיות כמו תעשייה מסורתית או אנרגיה, עקב מאפיינים שונים כמו צמיחה מהירה וחדשנות.
- שיטה לבדוק:
  - סיווג תעשיות: חלוקת המניות לפי תחומים או תעשיות.
  - השוואת תשואות: חישוב ממוצעים של תשואות לפי תחומים והשוואתן כדי לראות אם יש הבדל משמעותי בתשואות בין תחומים שונים.

## שיטות מחקר

### 1. ניתוח נתונים תיאורטי:

- חישוב ממוצעים וסטיות תקן: ניתוח התשואות היומיות והחודשיות של מניות ומדדים והסקת מסקנות לגבי התפלגותן.
- גרפים וויזואליזציות: הצגת גרפים של תשואות, מחירים ותנודתיות כדי להבין את ההתנהגות והקשרים בין המניות.

### 2. ניתוח סטטיסטי מתקדם:

- קורלציה ורגרסיה: בדיקת הקשרים בין תשואות המניות לבין ביצועי המדדים בעזרת ניתוח קורלציה ורגרסיה.
- מבחני השערות: ביצוע מבחנים סטטיסטיים כדי לאמת את השערות המחקר.

### 3. השוואת תחומים ותעשיות:

- הבדל בין תעשיות: השוואת תשואות בין מניות בתעשיות שונות כדי לזהות אם יש הבדל משמעותי בתשואות על פי תחום עיסוק.

באמצעות השיטות הנ"ל, ניתן לחקור ולהבין את התנהגות השוק, תשואות המניות, והשפעתם של גורמים כלכליים על הביצועים הפיננסיים.

## נתונים:

הנה קישור לתיקיית ההנתונים בענן:

<https://drive.google.com/drive/folders/1w9qGOtTrbuugDNxW06oBdSOkCg5v1HUx?usp=sharing>

תיאור הדאטה יחד עם סטטיסטיקה תיאורית פשוטה (יש לנו 3 קבצים שמכילים נתונים משנת 2014 - 2024):

1. נתוני מניות, הקובץ נקרא: sp500\_stocks

- מספר רשומות: 1,860,094

- עמודות:

- o Date תאריך (פורמט אובייקט)

- o Symbol סימול המניה (מחרוזת)

- o Adj Close מחיר סגירה מותאם (float64)

- o Close מחיר סגירה (float64)

- o High מחיר מקסימום (float64)

- o Low מחיר מינימום (float64)

- o Open מחיר פתיחה (float64)

- o Volume נפח מסחר (float64)

סטטיסטיקה תיאורית:

מאפיין	ממוצע	סטיית תקן	מינימום	מקסימום
מחיר סגירה מותאם (Adj Close)	96.05	207.49	0.20	9,321.27
מחיר סגירה (Close)	103.91	207.80	0.22	9,321.27
נפח מסחר (Volume)	6,961,111	33,132,470	0	3,692,928,000

2. נתוני מדד sp500\_index:

- מספר רשומות: 2,516

- עמודות:

- o Date תאריך (פורמט אובייקט)

- o S&P 500 ערך מדד (float64) S&P 500

3. נתוני חברות sp500\_companies:

- מספר רשומות: 503

• עמודות:

- Exchange בורסה (מחרוזת) -
- Symbol סימול המניה (מחרוזת) -
- Shortname שם קצר (מחרוזת) -
- Longname שם מלא (מחרוזת) -
- Sector מגזר (מחרוזת) -
- Industry תעשייה (מחרוזת) -
- Currentprice מחיר נוכחי(float64) -
- Marketcap שווי שוק(int64) -
- Ebitda - EBITDA (float64) -
- Revenuegrowth צמיחה בהכנסות(float64) -
- City עיר (מחרוזת) -
- State מדינה (מחרוזת) -
- Country מדינה (מחרוזת) -
- Fulltimeemployees עובדים במשרה מלאה(float64) -
- Longbusinesssummary סיכום עסקי (מחרוזת) -
- Weight משקל(float64) -

סטטיסטיקה תיאורית:

• מחיר נוכחי:(Currentprice)

- ממוצע: 104.91
- סטיית תקן: 108.83
- מינימום: 1.10
- מקסימום: 1,491.00

• שווי שוק:(Marketcap)

- ממוצע: 272,000,000,000
- סטיית תקן: 374,000,000,000
- מינימום: 7,000,000,000
- מקסימום: 2,700,000,000,000



## מניפולציות:

1. טעינת הדאטה, טעינת קבצים: הקבצים נטענו כדאטה פריימס באמצעות `pd.read_csv()`.

2. ניקוי נתונים

- טיפול בערכים חסרים:

- נבדקו ערכים חסרים בכל דאטה פריימ באמצעות `isna().sum()`.
- עבור `stocks_df`: שימוש בשיטה `forward fill` כדי למלא ערכים חסרים עם הערך הקודם בעמודה.
- עבור `companies_df`: מילוי ערכים חסרים בעמודות קטגוריאליות עם ערך 'Unknown' ובעמודות מספריות עם הממוצע או החציון.
- עבור `index_df`: לא היו ערכים ריקים.
- מילוי ערכים חסרים בעמודות מספריות באמצעות `KNN Imputer`, שמתחשב בשכנים הקרובים כדי למלא את הערכים החסרים.
- הדפסת ערכים חסרים לפני ואחרי הטיפול בהם.

- הסרת ערכים כפולים:

- הוסרו ערכים כפולים באמצעות `drop_duplicates()`.

3. המרת סוגי נתונים

- המרת תאריכים:

- עמודת `Date` הוסבה לפורמט תאריך באמצעות `pd.to_datetime()` תוך שימוש בפורמט נכון לשימוש.
- תיקון שמות עמודות:
  - הסרת רווחים ושינוי שמות העמודות לאותיות קטנות עם קווים תחתונים במקום רווחים, כדי להקל על השימוש בעמודות, השתמשתי: `str.strip().str.lower().str.replace(' ', '_')`

4. נרמול של עמודות מספריות:

- שימוש ב-`MinMaxScaler` כדי לשריין את העמודות המספריות של מחירי המניות כך שיתאימו לתחום בין 0 ל-1.

5. מיזוג הנתונים (מיזוג `DataFrames`):

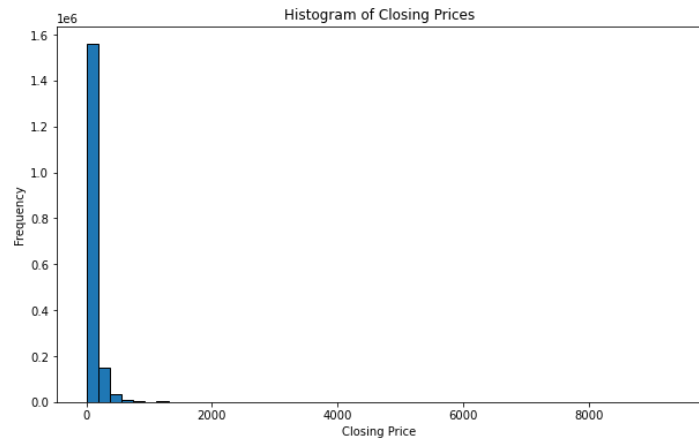
- מיזוג `stocks_df` ו-`index_df` על פי עמודת התאריך.
  - מיזוג עם `companies_df` על פי עמודת הסמל (`symbol`).
- ביצעתי ניקוי יסודי של הדאטה על מנת להבטיח שהנתונים יהיו נקיים ומוכנים לניתוח. התייחסות לערכים חסרים, הסרת כפילויות, תיקון סוגי נתונים וניהול שגיאות עזרו בהכנת הדאטה לצורך ניתוח מתקדם יותר.

## דיאגרמות פשוטות:

דיסטוגרמה: התפלגות מוטה של מחירי הסגירה.

תובנות:

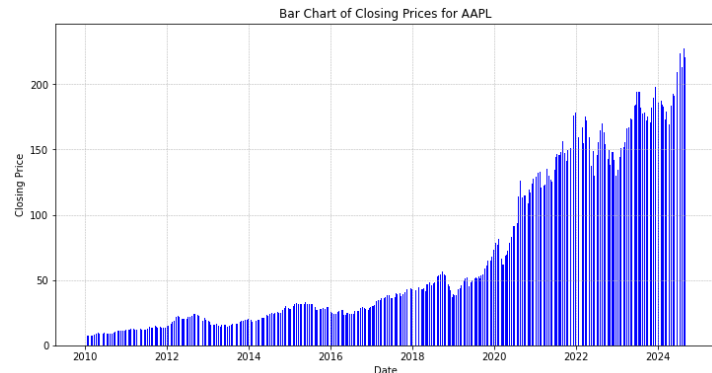
1. רוב המחירים נמוכים – ריכוז גבוה מאוד של מחירים נמוכים, קרוב לאפס.
2. מעט מחירים גבוהים – יש מספר קטן של מניות עם מחירים גבוהים מאוד (עד כ-9,000).
3. התפלגות מוטה – ההתפלגות מוטה ימינה, כלומר רוב המניות במחירים נמוכים, עם זנב ארוך של מחירים גבוהים.



תרשים עמודות מציג מחירי הסגירה של מניית Apple (AAPL) לאורך זמן מ-2010 ועד 2024

תובנות:

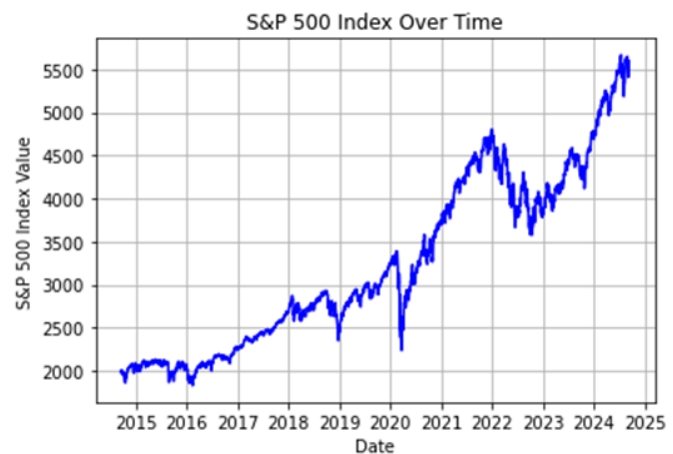
הגרף מציג עלייה חדה במחירי מניית Apple מאז 2010, עם צמיחה בולטת לאחר 2019. ניתן לראות תנודתיות גבוהה בעיקר בשנים 2020-2022, והמחירים מגיעים לשיאים חדשים עד 2024.



גרף קו מציג את ערכי מדד ה S&P 500 במהלך תקופה של מספר שנים (2015-2024).

תובנות:

מדד ה S&P 500 מראה עליות ארוכות טווח עם תנודתיות מסוימת, במיוחד במהלך משברים כלכליים.



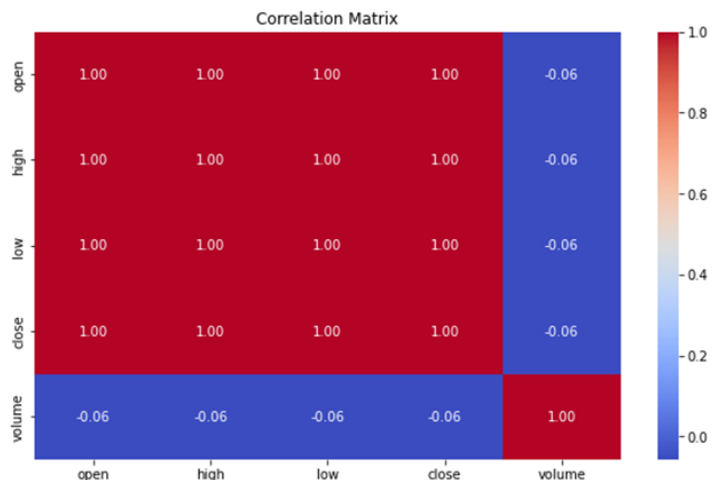
מטריצת קורלציה מראה את הקשרים בין התכונות השונות של הנתונים.

תוכנות:

מתאם גבוה בין מחירים: open, high, low ו-close מתואמים באופן מלא (1), מה שמעיד על תנועה זהה של המחירים.

מתאם חלש עם מחזור המסחר: מתאם חלש מאוד (-0.06) בין מחירי המניות למחזור המסחר (volume), מה שמרמז שאין קשר חזק בין השניים.

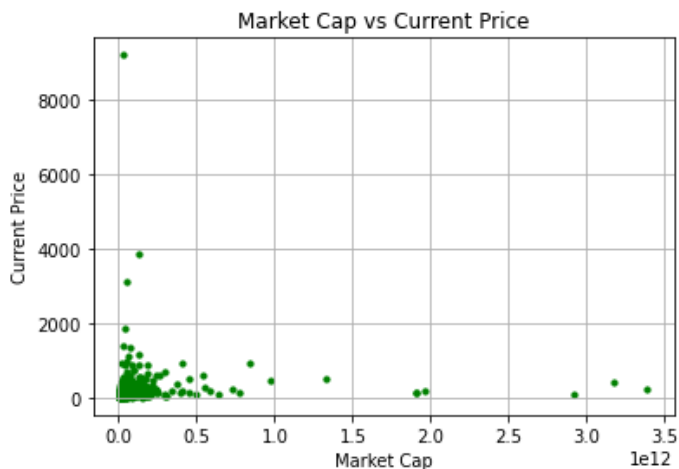
מחזור המסחר לא מתואם עם המחירים: מחזור המסחר מראה מתאם חזק רק עם עצמו ולא עם המחירים.



הקשר בין ערך שוק של חברות (Market Cap) לבין מחיר מניה נוכחי (Current Price) בחברות הנסחרות ב-S&P 500.

תוכנות:

- רוב החברות ב-S&P 500 הן בעלות מחירים סבירים למניה, אך יש חריגות של מחירי מניות גבוהים במיוחד.
- אין בהכרח קשר לינארי ברור בין ערך שוק למחיר המניה, וחברות גדולות מאוד לא בהכרח מתאפיינות במחירי מניות גבוהים.



## עיבוד הנתונים:

**השערה 1:** תשואות יומיות של מניות שונות יהיו שונות זו מזו, כאשר מניות של חברות שונות מציגות רמות שונות של תשואות יומיות עקב הבדלים בעסקי החברות, מגמות השוק, ותנודתיות.

- חישוב תשואות יומיות: חישוב תשואות יומיות לכל מניה על פי שינוי במחיר הסגירה.
- השוואת תשואות: השוואת התשואות היומיות של מניות שונות באמצעות סטטיסטיקות תיאוריות כמו ממוצעים וסטיית תקן, והשוואה בין התשואות המתקבלות עבור מניות שונות.

בקוד שביצעת, נכללו השלבים הבאים:

1. חישוב תשואות יומיות: תחילה חישוב את התשואות היומיות (daily returns) עבור כל אחת מהמניות המוכרות ביותר. התשואה היומית מחושבת כשינוי אחוזי בין מחירי הסגירה של מניה בין יום ליום.

2. גרף תשואות יומיות: נוצר גרף שמציג את התשואות היומיות של החברות הנבחרות (Apple, Microsoft, Google, Amazon, Tesla).

- ציר X מייצג את השנים (הצגת ימים או חודשים תגרום לגרף להיות מאוד עמוס ולא קריא)

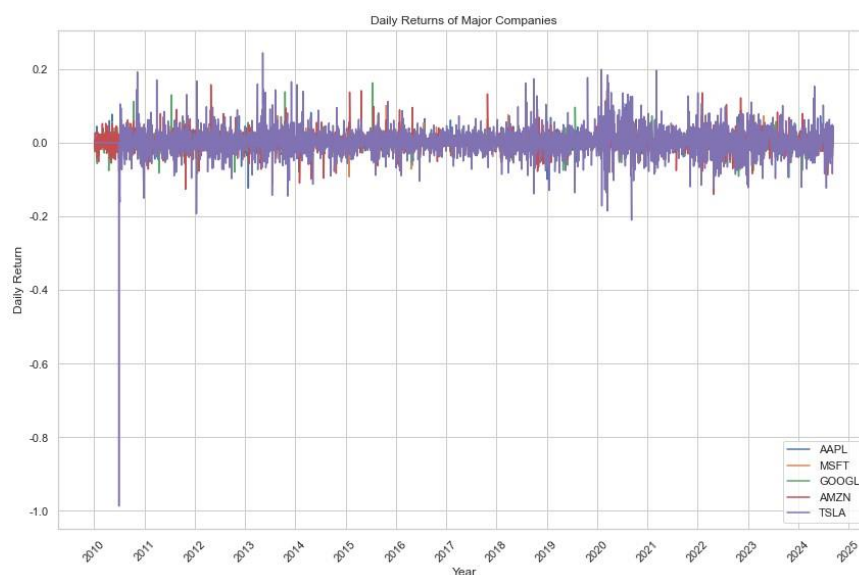
- ציר Y מציג את התשואה היומית (שהיא אחוז השינוי במחיר הסגירה).

3. אגדה: הקוד יצר אגדה (legend) שמסבירה איזה צבע מייצג כל חברה. זה מאפשר זיהוי ברור של קו התשואות עבור כל מניה בנפרד.

4. ממוצעים וסטיית תקן: הקוד גם מחשב את ממוצע התשואות היומיות ואת סטיית התקן של התשואות עבור כל חברה, כדי שניתן יהיה להשוות את הביצועים בין החברות על בסיס סטטיסטי.

בקצרה:

הקוד מבצע חישוב של תשואות יומיות, יוצר גרף שמציג את התשואות היומיות של מספר חברות טכנולוגיה, ומוסיף אגדה, והנה הגרף שנוצר:



## תובנות על הגרף:

התרשים מציג את התשואות היומיות של חברות טכנולוגיה מרכזיות, כולל AAPL (Apple), MSFT (Microsoft), GOOGL (Google), AMZN (Amazon) ו-TSLA (Tesla) לאורך השנים משנת 2010 ועד 2024.

הסברים עיקריים:

1. ציר X (זמן): ציר הזמן מציג את השנים בין 2010 ל-2024. כל נקודה על הציר מייצגת את התאריך שבו חישוב את התשואה היומית.

2. ציר Y (תשואה יומית): ציר ה-Y מציג את התשואה היומית באחוזים (בדמות חלקים עשרוניים), כלומר השינוי באחוזים במחיר הסגירה של המניות יום לאחר יום. תשואה חיובית משמעה עלייה בערך המניה, ותשואה שלילית משמעה ירידה.

3. חברות שונות (אגדים שונים): כל צבע מייצג חברה שונה (לדוגמה, כחול = Apple, אדום = Microsoft). כל קו מייצג את התשואות היומיות של החברה.

מאפיינים:

- תנודתיות גבוהה בתחילת התקופה (2010): ניתן להבחין בקפיצות גדולות בתשואות בתחילת התקופה, במיוחד ב-2010, שם ישנם תנודות חדות מאוד למטה, המגיעות עד ל-100% (השפעות אפשריות של משבר כלכלי או אירועים יוצאי דופן).

- מגמה יציבה יותר בהמשך: ככל שמתקדמים קדימה בשנים, התנודתיות נראית מתונה יותר, כלומר השינויים היומיים הם קטנים יותר.

- תנודות סביב אירועים משמעותיים: ניתן לראות "קפיצות" גבוהות יותר בשנת 2020 (בזמן תחילת מגפת הקורונה) ובשנת 2022, שמייצגות תקופות של חוסר וודאות כלכלית או משברים שיכלו להשפיע על שוק המניות.

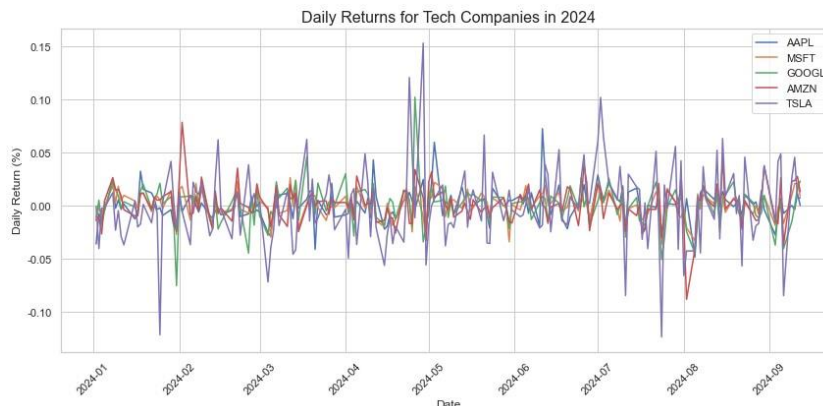
סיכום:

- התשואות היומיות משתנות בצורה דרמטית בהתאם לאירועים כלכליים גדולים ומשברים עולמיים, אך יש גם מגמות של יציבות במהלך תקופות שקטות יותר.
- הגרף מאפשר לראות את ההבדלים בתשואות בין החברות השונות, כאשר חברות טכנולוגיה כמו Apple, Microsoft ו-Tesla מציגות תנודתיות בולטת במהלך העשור האחרון.

## קישור הגרף להשערת המחקר:

- תשואות יומיות של מניות שונות יהיו שונות זו מזו: מהגרף ניתן לראות שהתשואות היומיות של מניות שונות אכן מציגות שונות מסוימת זו מזו. ישנן תקופות בהן חברה מסוימת חווה תשואות גבוהות או נמוכות יותר בהשוואה לאחרות, מה שממחיש את ההבדלים ברמות התשואות היומיות בין המניות השונות.
- מניות של חברות שונות מציגות רמות שונות של תשואות עקב הבדלים בעסקי החברות, מגמות השוק ותנודתיות: ההבדלים בתשואות ובתנודתיות שנראים על פני הגרף יכולים לשקף את ההשפעות השונות של המגמות הכלכליות והפיננסיות על כל חברה. לדוגמה, חברות טכנולוגיה כמו Apple ו-Tesla עשויות להיות רגישות יותר לאירועים מסוימים בתעשייה, מה שיכול ליצור שונות בתשואות שלהן לעומת חברות אחרות.

ביצעתי אותם שלבים עם פילטר לשנת 2024 כדי להראות בצורה ימית וברורה יותר את השינוי בתקופה קצרה.  
 בקוד: נבחרה תקופה שמתמקדת רק בשנת 2024, כלומר כל התשואות והנתונים מוצגים רק עבור השנה הזו. הסינון  
 בוצע על עמודת התאריכים, כדי לוודא שרק נתונים מהשנה הנוכחית (2024) נכללים בניתוח.



סיכום:

תנודתיות: סטיית התקן של התשואות היומיות יכולה להצביע על רמות התנודתיות השונות בין המניות, וגם זה מתיישב עם ההשערה שמניות שונות מציגות תשואות שונות עקב מאפיינים שונים של החברות.

לכן, התוצאה תומכת בהשערה שלך.

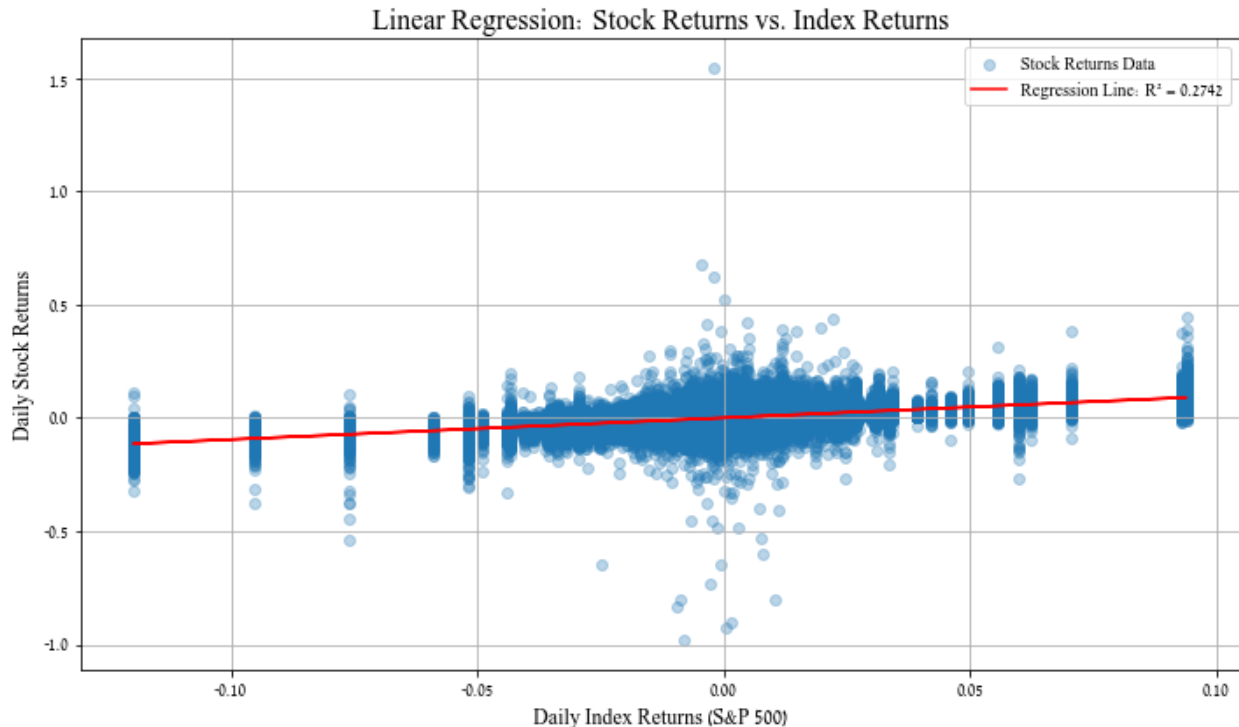
**השערה 2:** יש קשר בין תשואות מניות לבין ביצועי מדדים כמו S&P 500

○ השערה: תשואות המניות יהיו מקושרות לביצועים של מדדים רחבים כמו ה S&P 500-כאשר שינוי במדד ייתכן ויגרום לשינויים גם בתשואות המניות.

○ שיטה לבדוק:

- תשואות חודשיות של מניות ומדדים: חישוב תשואות חודשיות עבור מניות ונתוני מדד.
- קשר בין משתנים: ביצוע ניתוח קורלציה או רגרסיה כדי לבדוק את הקשר בין תשואות מניות לתשואות המדד לאורך זמן.

הקשר בין תשואות המניות למדד ה-S&P 500: ניתוח סטטיסטי מתקדם:



הקורלציה והמשוואה של הרגרסיה מספקות כמה תובנות חשובות על הקשר בין תשואות המניות למדד ה-S&P 500:

### 1. קורלציה (0.5237):

- הקורלציה בין התשואות של המניות לבין התשואות של מדד ה-S&P 500 היא 0.5237, כלומר קיימת קורלציה חיובית בינונית בין שתי התשואות.
- ערך חיובי מעל 0.5 מראה שיש קשר חיובי בין ביצועי המדד וביצועי המניות: כאשר המדד עולה, יש סיכוי טוב שגם המניות יעלו, ולהיפך.
- עם זאת, מכיוון שהערך רחוק מ-1, הוא לא מצביע על קשר חזק מאוד. כלומר, יש משתנים נוספים שמשפיעים על תשואות המניות מעבר למדד.

### 2. מקדם ה- $R^2$ (0.2742):

- ערך ה- $R^2$  (מקדם הדטרמינציה) של 0.2742 אומר ש-27.42% מהשונות בתשואות המניות מוסברת על ידי התשואות של מדד ה-S&P 500.
- המשמעות היא שיותר מ-70% מהשונות בתשואות המניות נגרמות על ידי גורמים אחרים שלא מוסברים רק דרך התשואות של המדד.

### 3. משוואת הרגרסיה:

- המשמעות היא שכל עלייה של 1% בתשואות המדד (S&P 500) מובילה בממוצע לעלייה של 0.9551% בתשואות המניות.
- האינטרספט (0.0001) הוא זניח יחסית, כלומר במצב שבו התשואות של המדד אפס, המניות צפויות להישאר באותו מצב בממוצע.

## סיכום:

- יש קשר חיובי בינוני בין תשואות המניות לתשואות המדד, אך הוא לא מושלם. מדד ה-S&P 500 מסביר חלק מהשונות בתשואות המניות, אך ישנם גורמים נוספים שמשפיעים.
- משוואת הרגרסיה מסבירה את הקשר הלינארי בין התשואות, כאשר עלייה במדד מובילה כמעט לאותו שיעור עלייה במניות.

הניתוח הסטטיסטי שבוצע מספק תובנות התומכות בהשערה 2: יש קשר בין תשואות מניות לבין ביצועי מדדים כמו S&P 500. הנה איך הממצאים מתחברים להשערה:

1. קורלציה חיובית (0.5237):

- הקורלציה החיובית בין תשואות המניות לתשואות מדד ה-S&P 500 תומכת בהשערה שיש קשר חיובי בין התשואות.
- אמנם הקורלציה אינה חזקה מאוד (היא לא מתקרבת לערך של 1), אבל היא עדיין מראה שיש קשר מובהק ביניהם. ככל שהמדד עולה, תשואות המניות נוטות לעלות גם כן, ולהיפך.

2. רגרסיה ( $R^2 = 0.2742$ ):

- ערך ה- $R^2$  (27.42%) מצביע על כך שכמעט רבע מהשונות בתשואות המניות מוסבר על ידי התשואות של מדד ה-S&P 500.
- זה מעיד שהמדד מהווה גורם חשוב שמשפיע על תשואות המניות, אך ישנם גורמים אחרים שמסבירים את השאר, כגון גורמים ספציפיים לחברה, אירועים מאקרו-כלכליים אחרים, או תעשיות שונות.

3. משוואת הרגרסיה:

- המשוואה ( $\text{Stock Returns} = 0.9551 * \text{Index Returns} + 0.0001$ ) מראה שהמדד משפיע בצורה כמעט פרופורציונלית על המניות (כמעט 1:1).
- זה מצביע על כך שביצועי המדד הם אינדיקטור טוב לביצועי המניות, אך לא הגורם הבלעדי.

מסקנה לגבי ההשערה 2:

השערה 2 מאוששת באופן חלקי. יש קשר חיובי בין ביצועי המניות והמדד, אך הוא לא מוחלט. ביצועי המדד מסבירים חלק משמעותי מהתשואות של המניות, אך לא את כולן, מה שמעיד שיש גם משתנים נוספים המשפיעים.

**השערה 3:** מניות בתחום טכנולוגיה מציגות תשואות גבוהות יותר מאשר מניות בתעשיות מסורתיות

- השערה: מניות בתחום הטכנולוגיה יציעו תשואות גבוהות יותר בהשוואה למניות בתעשיות כמו תעשייה מסורתית או אנרגיה, עקב מאפיינים שונים כמו צמיחה מהירה וחדשנות.
- שיטה לבדוק:

- סיווג תעשיות: חלוקת המניות לפי תחומים או תעשיות.
- השוואת תשואות: חישוב ממוצעים של תשואות לפי תחומים והשוואתן כדי לראות אם יש הבדל משמעותי בתשואות בין תחומים שונים.



## תיאור עיבוד הנתונים

### 1. סיווג תחומים

הגדרת סוגי תחומים: סיווגנו את התחומים ל"Technology" (טכנולוגיה) ו"Traditional" (מסורתי). תחומים שלא נחשבים לטכנולוגיה סווגו כ"Traditional".

### 2. סטטיסטיקות תשואות

חישוב סטטיסטיקות תשואות: חישבתי סטטיסטיקות לתשואות היומיות לכל סוג תחום (ממוצע, סטיית תקן, ערכים מינימליים ומקסימליים).

### 3. השוואת תשואות בין סוגי תחומים:

ממוצע תשואות: חישבנו את ממוצע התשואות, סטיית התקן, והקוורטילים (25%, 50%, 75%) לכל סוג תחום.

מבחן T: ביצענו מבחן T כדי לבדוק אם יש הבדל מובהק סטטיסטית בתשואות בין תחום הטכנולוגיה לתחומים המסורתיים. המבחן מבצע השוואה בין התשואות היומיות של שני הקטגוריות.

### 4. תוצאות עיבוד הנתונים:

סטטיסטיקות תשואות לפי סוג תחום

- Technology: תשואות עם ממוצע של 0.000849, סטיית תקן של 0.023252, עם טווח של ערכים בין -1.546093 ל-1.000000.
- Traditional: תשואות עם ממוצע של 0.000493, סטיית תקן של 0.020031, עם טווח של ערכים בין -0.985415 ל-0.745932.
- מבחן T:
- T-statistic: 6.4351
- P-value: 0.0000

המבחן מראה שיש הבדל מובהק סטטיסטית בתשואות בין תחום הטכנולוגיה לתחומים המסורתיים, כאשר p-value נמוך מ-0.05.

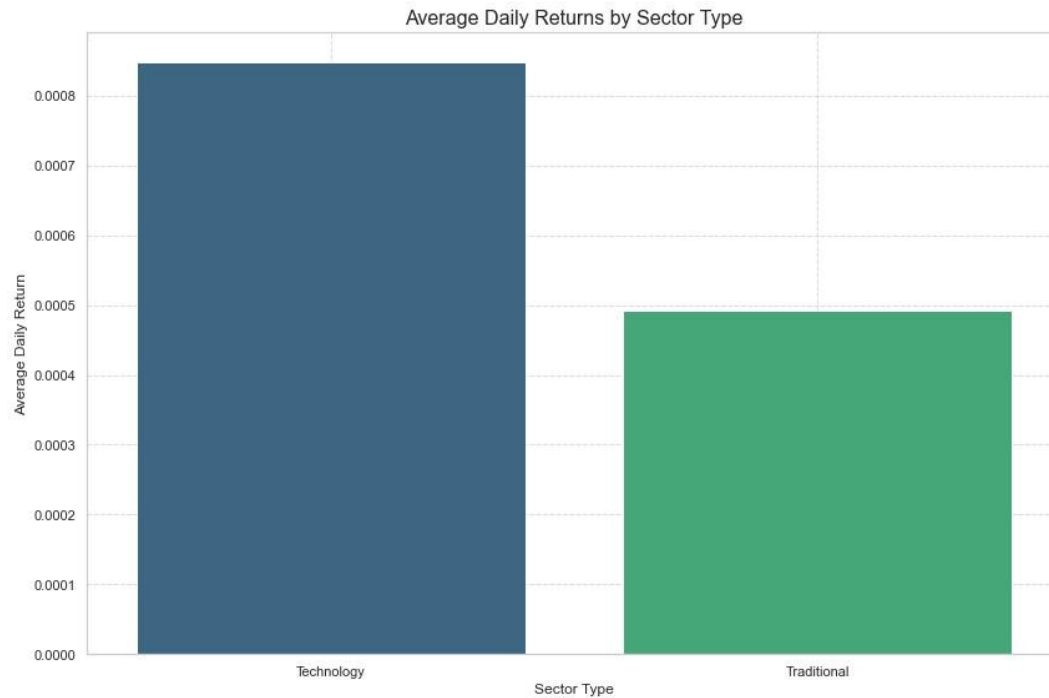
### 5. הגרפים

גרף עמודות: הצגנו גרף עמודות של התשואות היומיות הממוצעות לפי סוג תחום, כדי להמחיש את ההבדל בתשואות בין התחומים השונים.

### קשר להשערה 3:

השערה 3: מניות בתחום הטכנולוגיה מציגות תשואות גבוהות יותר מאשר מניות בתעשיות מסורתיות.

העיבודים והסטטיסטיקות שבוצעו, כולל המבחן T, נועדו לבדוק את השערה זו. התוצאות מראות הבדל מובהק סטטיסטית בתשואות היומיות בין מניות טכנולוגיה למניות בתעשיות מסורתיות, שתומכות בהשערה זו.



הגרף תומך בהשערה 3 בכך שהוא מציג בצורה ברורה שממוצע התשואות היומיות בסקטור הטכנולוגיה גבוה יותר מאשר בסקטור המסורתי. ניתן לראות שבטור של "Technology" התשואות היומיות גבוהות יותר בהשוואה ל"טור של "Traditional".

במילים אחרות, הגרף מציג תמיכה חזותית להשערה שמניות בתחום הטכנולוגיה מציגות ביצועים טובים יותר מבחינת תשואות יומיות בהשוואה למניות בתעשיות מסורתיות.

## חיזוי:

בחלק האחרון של הקוד בקובץ `final-project.py`, התוכנה חוזה את מחיר מניות החברות המוכרות בעולם שיש בסט הנתונים לפי שנה יחודש שהמשתמש מזין, הנה דוגמא:

```
Enter Year (2023-2030): 2026
Enter Month (1-12): 8
```



1. טעינת נתונים: הקוד מתחיל בטעינת נתוני מניות מקובץ CSV. הנתונים כוללים תאריכים, סימולים של מניות ומחירים מותאמים (`adj_close`).
  2. הכנת נתונים: המרתי את עמודות התאריכים לפורמט שניתן לעבוד איתו בקלות ומכין את הנתונים בצורה שתאים למודל התחזיות.
  3. אימון מודל: הקוד מאמן מודל תחזיות בשם Prophet על פי נתוני המניה שנבחרו. המודל לומד את דפוסי המחירים ההיסטוריים כדי לבצע תחזיות עתידיות.
  4. תחזיות עתידיות: לאחר אימון המודל, הוא יוצר תחזיות למחירים עתידיים עבור המניה.
  5. תחזיות עבור מספר מניות: הקוד מבצע את כל התהליך עבור מספר מניות שונות (כגון AAPL, MSFT, GOOGL, AMZN, TSLA) ומחשב את המחיר הממוצע המתחזה לכל מניה.
  6. הצגת תוצאות: בסוף, הקוד מציג את התחזיות באמצעות טבלה ומייצר גרף עמודות שמראה את מחירי התחזיות עבור כל מניה.
- באופן כללי, הקוד מספק תחזיות מחירים עתידיים עבור מניות שונות על בסיס נתונים היסטוריים, ומציג את התוצאות בצורה נוחה להבנה.

## דיון ומסקנות:

בהתבסס על ההשערות והמחקר שביצעתי, התוצאות מצביעות על כך שההשערות נבחנו והתקבלו מסקנות מעניינות. הנה כמה תובנות מהעבודה, וכיצד ההשערות עמדו במבחן:

השערה 1: תשואות יומיות משתנות בין מניות שונות

תוצאות:

- התשואות היומיות של מניות שונות אכן משתנות זו מזו, כפי שניתן לראות מהתנודתיות השונה בין חברות כמו Tesla, Apple, Microsoft, Google, Amazon.
- חברות טכנולוגיה מרכזיות מציגות רמות שונות של תנודתיות ותשואות יומיות.
- הקוד והגרף מדגימים את ההבדלים, וממוצעי התשואות וסטיות התקן משקפים את השונות בין המניות.

האם ההשערה עמדה במבחן?

כן, ההשערה עברה את המבחן. הנתונים מראים הבדלים מובהקים בתשואות היומיות בין המניות.

השערה 2: קיימת קורלציה חיובית בין תשואות מניות לבין מדד ה-S&P 500

תוצאות:

- הקורלציה בין תשואות המניות לבין מדד ה-S&P 500 היא 0.5237, מה שמצביע על קשר חיובי בינוני בין השניים.
- ערך ה- $R^2$  של 0.2742 מראה כי כ-27.42% מהשונות בתשואות המניות מוסברת על ידי תשואות המדד.
- משוואת הרגרסיה ( $\text{Index Returns} + 0.0001 * 0.9551$ ) מציעה כי עלייה של 1% במדד ה-S&P 500 תוביל בממוצע לעלייה של כ-0.9551% בתשואות המניות.

האם ההשערה עמדה במבחן?

כן, ההשערה עברה את המבחן באופן חלקי. הנתונים מראים שקיים קשר חיובי בין תשואות המניות וביצועי מדד ה-S&P 500, אך הוא אינו חזק במיוחד וישנם גורמים נוספים המשפיעים על תשואות המניות.

השערה 3: מניות בתחום הטכנולוגיה מציגות תשואות גבוהות יותר מאשר מניות בתעשיות מסורתיות

תוצאות:

- Technology: ממוצע תשואות של 0.000849, סטיית תקן של 0.023252, עם טווח ערכים בין -1.546093 ל-1.000000.
- Traditional: ממוצע תשואות של 0.000493, סטיית תקן של 0.020031, עם טווח ערכים בין -0.985415 ל-0.745932.
- מבחן T:  $T\text{-statistic} = 6.4351$ ,  $P\text{-value} = 0.0000$ . מה שמעיד על הבדל מובהק סטטיסטית בין התשואות של שני התחומים.

האם ההשערה עמדה במבחן?

כן, ההשערה עברה את המבחן. התשואות של מניות הטכנולוגיה גבוהות יותר באופן מובהק בהשוואה למניות בתעשיות המסורתיות, כפי שמוכח על ידי ממוצעי התשואות וההבדלים המובהקים מבחינה סטטיסטית.

## הרחבות לעבודה

1. השוואת תשואות לאורך זמן עבור סקטורים שונים:

ניתן להרחיב את הניתוח על ידי בחינת תשואות לאורך פרקי זמן ארוכים יותר עבור סקטורים שונים (לא רק טכנולוגיה ומסורת) כדי לזהות מגמות לאורך השנים. למשל, האם יש מגמה מתמשכת של עליונות בתחום הטכנולוגיה או האם ישנם סקטורים נוספים שמציגים ביצועים חזקים בתקופות כלכליות שונות?

2. ניתוח רגישות למאורעות כלכליים:

ניתן לנתח כיצד סקטורים שונים מושפעים מאירועים כלכליים גלובליים כמו משברים כלכליים, רפורמות, שינויים במדיניות המוניטרית, או תהליכים פוליטיים. אירועים כגון מגפת הקורונה או משברים פיננסיים יכולים לחשוף הבדלים נוספים בין תעשיות שונות.

3. השוואה מול מדדים נוספים:

ניתן להוסיף לניתוח מדדים נוספים מעבר ל-S&P 500. ולהשוות את השפעתם על התשואות של מניות הטכנולוגיה והמניות המסורתיות.

4. מודלים מתקדמים לחיזוי תשואות:

שילוב מודלים מתקדמים לחיזוי תשואות, כגון מודלים של למידת מכונה (כמו Random Forest או Gradient Boosting), יכול לאפשר חיזוי מדויק יותר של התשואות בהתבסס על משתנים נוספים, כגון אינדיקטורים מאקרו-כלכליים, אירועים חיצוניים, או נתונים פנימיים של החברות.

5. השוואת סקטורים גאוגרפיים:

ניתן להרחיב את הניתוח להשוואת סקטורים גאוגרפיים, לדוגמה, להשוות את ביצועי מניות הטכנולוגיה בארה"ב מול מדינות אחרות, כמו אירופה או אסיה, כדי להבין את ההשפעות הגלובליות השונות על ביצועי המניות.

## סיכום

במהלך העבודה נותחו התשואות היומיות של מניות ממגזרים שונים תוך השוואה למדד ה-S&P 500, במטרה לבדוק את הקשר ביניהם ולבחון האם תחום הטכנולוגיה מציג תשואות גבוהות יותר ממניות מסורתיות.

תוצאות הניתוחים העלו כמה ממצאים מרכזיים:

1. השערה 1: אוששה. נמצא כי התשואות היומיות משתנות באופן מובהק בין מניות שונות, עם הבדלים ניכרים בתנודתיות בין חברות טכנולוגיה מובילות כמו Google, Microsoft, Apple, ו-Tesla.
2. השערה 2: אוששה באופן חלקי. הקורלציה בין תשואות המניות למדד ה-S&P 500 היא חיובית, אך לא חזקה. מדד ה-S&P 500 מסביר כ-27% מהשונות בתשואות המניות, מה שמעיד על כך שגורמים נוספים משפיעים על ביצועי המניות.
3. השערה 3: אוששה. נמצא הבדל מובהק סטטיסטית בין התשואות היומיות של מניות טכנולוגיה לתשואות של מניות בתעשיות מסורתיות, כאשר מניות טכנולוגיה מציגות ביצועים גבוהים יותר באופן מובהק.